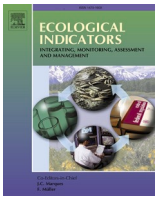
Ecological Indicators 131 (2021) 108200

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind



Original Articles

Evaluation of the factors explaining the use of agricultural land: A machine learning and model-agnostic approach

Cláudia M. Viana [a,*], Maurício Santos [a], Dulce Freire [b], Patrícia Abrantes [a], Jorge Rocha [a]

[a] Centre for Geographical Studies, Institute of Geography and Spatial Planning, Universidade de Lisboa, Lisbon, Portugal
[b] Faculty of Economics, University of Coimbra, Coimbra, Portugal

ARTICLE INFO

Keywords:
Cropland
Interpretability
Artificial intelligence
xAI
LIME

ABSTRACT

To effectively plan and manage the use of agricultural land, it is crucial to identify and evaluate the multiple human and environmental factors that influence it. In this study, we propose a model framework to identify the factors potentially explaining the use of agricultural land for wheat, maize, and olive grove plantations at the regional level. By developing a machine-learning model coupled with a model-agnostic approach, we provide global and local interpretations of the most influential factors. We collected nearly 140 variables related to biophysical, bioclimatic, and agricultural socioeconomic conditions. Overall, the results indicated that biophysical and bioclimatic conditions were more influential than socioeconomic conditions. At the global interpretation level, the proposed model identified a strong contribution of conditions related to drainage density, slope, and soil type. In contrast, the local interpretation level indicated that socioeconomic conditions such as the degree of mechanisation could be influential in specific parcels of wheat. As demonstrated, the proposed analytical approach has the potential to serve as a decision-making tool instrument to better plan and control the use of agricultural land.

1. Introduction

The current global trends of population growth, accelerated urbanisation, and environmental changes, which are associated with the encroachment of agricultural land (Foley et al., 2011; Radwan et al., 2019), agricultural land abandonment (Castillo et al., 2021), and agricultural land fragmentation (Gomes et al., 2019; Postek et al., 2019), have an influence on food production and food security (Godfray et al., 2010; Wu et al., 2014). For the coming decades, enhancing and maintaining food supply will require the efficient use of agricultural land (FAO, 2017; Wu et al., 2014). However, multiple factors (e.g. natural and environmental) that vary both temporally and spatially determine and affect the use of agricultural land (Akpoti et al., 2019; Lambin et al., 2001; Ndamani and Watanabe, 2017). Thus, more studies are needed to better identify and evaluate the factors influencing agricultural land use under different cross-scale and geographical contexts.

Traditionally, empirical and conventional statistical methods such as principal component analysis (PCA), clustering methods, regression, and other linear approaches have been used to better understand the factors influencing land use (Braimoh, 2009; Marcos-Martinez et al., 2017; Santiphop et al., 2012; Velásquez-Milla et al., 2011). While the application of such methods can provide useful information to support effective planning and management measures as well as better-informed decisions concerning efficient land use, they present some analytical limitations. For instance, these statistical methods may not fully capture nonlinear behaviour or discard the effects of heterogeneity and spatial autocorrelation from the analysis (Cartone and Postiglione, 2020; Demšar et al., 2013; Jombart et al., 2008).

Conversely, machine learning (ML), which is a subfield of artificial intelligence (AI), has successfully overcome the limitations of statistical methods. Compared to traditional methods, ML is recognised to achieve superior or at least equivalent accuracy outcomes (Lima et al., 2015; Ren et al., 2020; Shortridge et al., 2016). In turn, ML approaches have many advantages, such as the capability to deal with data of different types, structures, and quantities (i.e. big data) (Molnar, 2019), being non-sensitive to the scale of variables (meaning there is no need for variable normalisation); therefore, it is possible to exploit and combine different data resources to model complex nonlinear relationships that describe agricultural land-use systems.

Owing to the great variety of robust algorithms and flexible model

* Corresponding author.
E-mail address: claudiaviana@campus.ul.pt (C.M. Viana).

https://doi.org/10.1016/j.ecolind.2021.108200
Received 15 June 2021; Received in revised form 6 September 2021; Accepted 10 September 2021
Available online 14 September 2021

structures (e.g. artificial neural networks (ANNs) and random forests (RFs)), ML models represent a potential solution to the requirements of different land-use modelling applications (Hagenauer et al., 2019). Despite the potential advantages, ML algorithms remain mostly under a 'black box' formulation, which means that without further intervention, it is not possible to directly interpret or retrace how a model performs inference or prediction owing to the many internal weights or structural information (Molnar, 2019). Nevertheless, the explainable AI (xAI) has recently emerged as an important research area, which proposes advanced statistical measures and visualisation tools to enhance the interpretability of ML (Carvalho et al., 2019; Molnar, 2019; Murdoch et al., 2019). For instance, post-hoc techniques such as model-agnostic models have been proposed as interpretability methods to provide explanations about the function underlying the general behaviour of ML models (Molnar, 2019; Murdoch et al., 2019; Ribeiro et al., 2016a). The main advantage of a model-agnostic approach is its flexibility as it can deal with the opacity of any kind of black box ML model and gather interpretability, which is a critical aspect when the ML model outcomes are used as a basis for decision making (Ribeiro et al., 2016a). Currently, there are some examples of model-agnostic interpretation methods that are global or local in scope, such as the permutation feature importance (PFI), partial dependence plots (PDPs), or local surrogate models (e.g. local interpretable model-agnostic explanations (LIME)) (Molnar, 2019).

To date, ML models have been successfully applied in a wide range of Earth and environmental science studies, for example, in estimating air pollution (Ren et al., 2020), predicting dengue importation (Salami et al., 2020), modelling coastal fish communities (Lehikoinen et al., 2019), and predicting marine fish distributions (Zhang et al., 2019). In the scientific field of land-use modelling, ML has been mostly used for image classification and land use/land cover (LULC) mapping (Abdi, 2020; Raczko and Zagajewski, 2017), as well as to simulate future LULC changes (Gomes et al., 2019; Hagenauer et al., 2019). However, the recently developed areas of xAI research and model-agnostic methods, which provide the ML model interpretability needed to enhance scientific consistency (Molnar, 2019), have rarely been introduced to agricultural land modelling studies.

Accordingly, this study explores the use of an ML model coupled with a model-agnostic approach to increase the understanding of human and environmental factors that can explain the use of agricultural land for three cropland plantations relevant to food security and Mediterranean basin ecosystems: wheat, maize, and olive groves (FAO, 2018; Loumou and Giourga, 2003). Thus, we developed an analytical framework using the RF ML algorithm and PFI, PDPs, and LIME model-agnostic methods to provide global and local interpretabilities to understand how bioclimatic, biophysical, and socioeconomic conditions might explain the land used for these three cropland plantations. From a quantitative methodology perspective, this study demonstrates the usefulness of such methods to deal with some of the above-described analytical challenges, and provides novel insights into the use of agricultural land at the regional scale.

2. Material and methods

2.1. Case study and cropland context

The case study in which the modelling framework was developed is the Beja district located in southern Portugal, with an area of approximately 10,229.05 km^2 that covers 11% of Portugal's mainland territories (Fig. 1). In 2011, the district had a population of 152,758 inhabitants, distributed among 14 municipalities and 75 parishes (INE, 2012). The climate in Beja is influenced by its distance from the coast, with a Mediterranean climate characterised by hot and dry summers and wet and cold winters. This large predominantly agricultural region includes a vast landscape of intermingling cultures, such as wheat, olive groves, vineyards, and cork oak forests. In addition, in this region, we

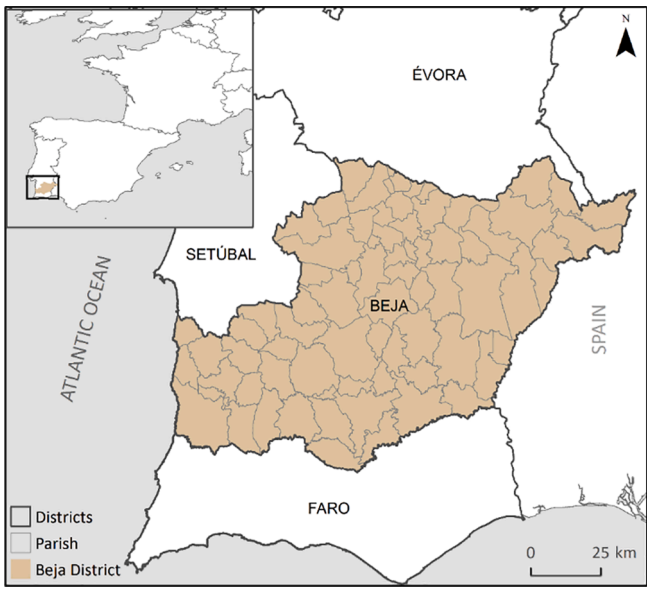


Fig. 1. Location of the Beja district in Portugal.

can find an agrosilvopastoral agricultural heritage system named Montado that has been indicated as a globally important agricultural system according to the Food and Agriculture Organization of the United Nations (FAO) (Correia, 1993; Koohafkan, 2016; Muñoz-Rojas et al., 2019). The region's high natural and economic value, in particular to food security and Mediterranean basin ecosystems, emphasizes the case study choice.

Over the last century, Beja consistently produced increasing amounts of wheat due to its ecological–biophysical conditions and parcel structures. The region is also characterised by an enlargement of olive grove plantations, whereby a change from open production to intensive and super-intensive production has been witnessed over the past decade due to the increased exploitation of water resources following the construction of Alqueva Dam (Viana et al., 2019; Viana and Rocha, 2020). Although maize plantations were historically confined to the northern regions of Portugal, which are more humid and have a higher water availability than the south, they can now be found in Beja as a result of the construction of irrigation systems during the past decades.

2.2. Experimental design

The model framework was developed to understand the multiple factors explaining the land used for wheat, maize, and olive grove plantations. The framework includes five main stages: (1) collection and pre-processing of spatial data, (2) data multicollinearity diagnosis, (3) ML model building, and (4) application of a model-agnostic approach for interpretability. The workflow of the process is shown in Fig. 2.

2.2.1. Data collection and pre-processing

2.2.1.1. Derivation of the response variable. The spatial locations of wheat, maize, and olive groves were obtained from the Portuguese Institute for Financing Agriculture and Fisheries (IFAP) (https://www.ifap.pt/isip/ows/). The IFAP provides vector 1:10,000 (polygons) structured data concerning the Land Parcel Identification System, which identifies the limit of parcels of national farming systems and classifies agricultural land use (reference data for 2020). This dataset is produced by the Portuguese government for the submission of applications for community aid and the execution of control actions for farmers. For analysis purposes, we generated random points inside the parcel features, with a minimum distance of 500 m between them, to avoid pseudo-replication and to increase the variance of the training data. A

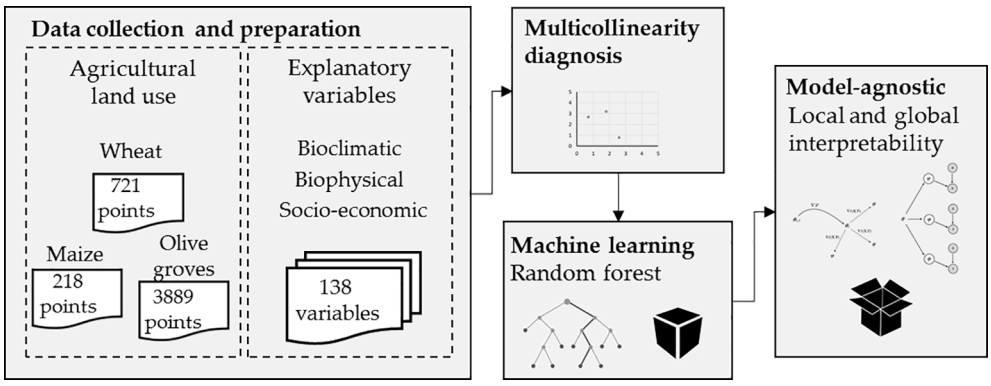


Fig. 2. Workflow of the modelling process.

total of 721, 218, and 3889 sample presence points were obtained for wheat, maize, and olive groves, respectively (Fig. 3). Thus, an equal number of real absences were selected and included in the final presence–absence file.

2.2.1.2. Derivation of the explanatory variables. Variables (factors) with probable explanatory relevance were selected based on relevant literature (e.g., Akpoti et al., 2019; Kourgialas, 2021; Li et al., 2018; Petit et al., 2011; Valayamkunnath et al., 2020); however, the initial dataset depended on data availability. Briefly, multiple variables were obtained and divided into (i) agricultural socioeconomic statistical data, which provide a current comprehensive information framework for the agricultural sector in the region; and (2) environmental data (bioclimatic and biophysical), which provide basic information on climate and the physical environment. The entire dataset encompassed 138 variables (both categorical and continuous). Table 1 presents a summary of the variables included in this study and their metadata (full data in Table A1 in the Appendix). It is worth noting that digital terrain model (DTM) data were used to compute the slope, and drainage network data were used to calculate the drainage density (km/km^2). In addition, land cover data and road networks were used to compute the Euclidian distance to waterbodies, urban areas, and roads. Socio-economic data was collected at the parish level and rasterized at a 100 m resolution. Therefore, all original data were resampled to a common spatial resolution of 100 m to match the climate data resolution and 1:25,000 variable minimum mapping unit (1 ha).

2.2.2. Modelling procedures

The first stage of our model approach encompassed a preliminary statistical analysis, during which we diagnostic the variables multicollinearity by calculating the variance inflation factor (VIF) (Dohoo et al., 1997; Lin, 2008). We calculated the VIF using the 'usdm' library in R statistical software, which excluded highly correlated variables from the initial set (n = 138) through a stepwise procedure (Naimi et al., 2014). Any variable with a VIF of > 5 was excluded from the model (James et al., 2013; Johnston et al., 2018). The VIF was calculated for each cropland individually and the final number of included variables was 42 for wheat, 25 for maize, and 44 for olive groves (see Table A2 in the Appendix).

The second stage involved building the ML model using the RF algorithm. Several studies have demonstrated that RF exhibits very similar performance or performs better than other ML algorithms (see Al-Fugara et al., 2020; Li et al., 2016; Wu et al., 2019; Yang et al., 2016). Furthermore, the RF algorithm has been described as robust in terms of fitting capacity during training and validation procedures, even with a small number of sample points (Luan et al., 2020; Moghaddam et al., 2020; Qi, 2012). We performed RF using the 'RandomForest' library in R statistical software (Liaw and Wiener, 2002) without pre-tuning the number of trees (n = 500) and setting the number of variables in the subset of each node to $\sqrt[2]{n}$ (Probst and Boulesteix, 2018). Each cropland was modelled separately. Due the RF model structures is important to tune and validate it. Therefore, we used k-fold cross-validation (k = 10) to enhance the model reliability and avoid overfitting (Meyer et al., 2018). We chose a data split of 70% for training and 30% for testing instead of an 80%–20% split due to the low number of presences for maize (2 1 8) and wheat (7 2 1) (Meyer et al., 2018).

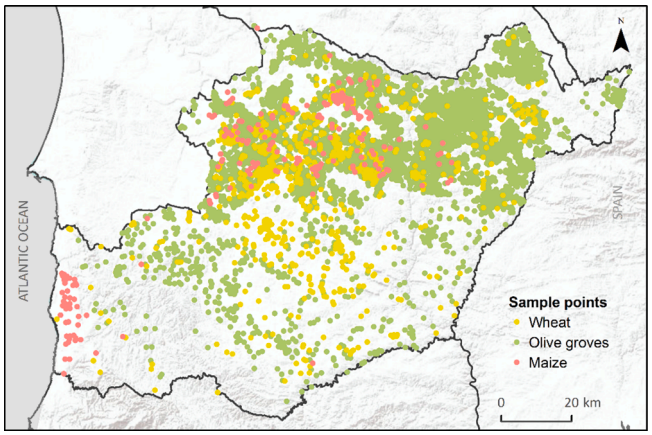


Fig. 3. Sample points spatial location. The sample size among the three crop plantations depicts the importance in terms of the territorial presence and spatial distribution of each crop plantation.

2.2.3. Model-agnostic approach for interpretability

The third stage involved interpretability of the ML model by using the model-agnostic approach to extract the post-hoc explanations. The model-agnostic approach provides an explanation based on the different behaviours of fitted complex models (e.g. RF algorithm), presenting information at the global level (i.e. what the model learned from the input variables) and at the local level (i.e. the rationales that the model provides for each estimate). Depending on the purpose of the analysis, different methods can be used jointly for global or local interpretability of the same model. In this study, we applied two methods for global interpretability and one method for local interpretability. The global interpretation was implemented using the following methods:

(a) The PFI method, which is commonly used to measure the increase in model performance error after a variable is permuted (i.e. randomly shuffled) (Molnar, 2019; Winkler et al., 2015). Specifically, this method allowed us to understand which variables contributed to the underlying ML model outcomes and quantify their importance scores. In this study, we computed both the area under the curve (AUC) and the R^2 value, and used the latter for the variable importance analysis.

Table 1
Summary of the variables included in the model.

Category	Code	Variable	Scale/Resolution	Year	Data Source
Socio-economic	V1-V55	Agricultural holdings (type of tenure, legal form, type of land use, livestock, with irrigable area)	1:25,000Parish statistical unit	2019	Statistics Portugal (2019)
	V56-V61	Agricultural types of machinery			
	V62	Familiar agricultural population			
	V63-V89	Sole agricultural holders (Age group, with 65 and more years old, female sex, level of education)			
	V90-V110	Utilised agricultural area			
	V111-V112	Irrigable area (ha) of agricultural holdings and AWU			
Bioclimatic	V113-V114	Mean temperature of the warmest and the coldest month of the year	100 *m*	1960–1990	Monteiro-Henriques et al. (2016)
	V115-V116	Annual positive temperature and positive precipitation			
	V117-V118	Mean maximum and minimum temperature of the coldest month			
	V119	Simple continentality index, or annual thermal amplitude			
	V120-V121	Thermicity índex and compensated thermicity índex			
	V122	Annual ombrothermic índex			
	V123	Ombrothermic index of the warmest bimonth of the summer quarter			
	V124-V125	Ombrothermic index of the summer quarter and the summer quarter plus the previous month			
	V126-V127	Positive precipitation (for dry and humid years)			
	V128-V129	Ombrothermic index (for dry and humid years)			
	V130-V131	Ombrothermic index anomaly (for dry and humid years)			
Biophysical	V132	Soil type	1:25,000	Static over time	DGADR (https://www.dgadr.gov.pt/)
	V133	Soil capacity			
	V134	Slope			IGEOE (http://www.igeoe.pt/cigeoesig/)
	V136	Drainage density			
	V137-V138-V135	Distance to urban, roads, and water bodies			DGT (2018) and OpenStreetMap (https://www.openstreetmap.org/)

(b) The PDPs method, which depicts the explanatory variables' overall relationship with the response variable (variable explanation probability) by imposing all occurrences to have the same variable value and measuring the marginal or average effect for this value on the model response (Apley and Zhu, 2016; Goldstein et al., 2013; Molnar, 2019). Therefore, it indicates the marginal effects of variables on the model outcomes, thereby identifying the threshold value at which these variables are likely to explain the land used for a specific cropland plantation.

In addition, local interpretation was implemented via:

(a) The LIME method, which trains a local surrogate model (a simple model such as a decision tree model) to reconstruct the inter logic workings around the individual observation (i.e. the local potential interactions occurring) (Ribeiro et al., 2016b). Formally, the LIME interpretability constraint is defined as:

$$explanation(x) = argmin_{g \in G} L(f, g, \pi x) + \Omega(g) \quad (1)$$

where the explanation model (x) is the model(g), for example, a generalised linear model, which minimises the loss (L) using the mean squared error, which measures how close the explanation is to the prediction of the original model (f), for example, a RF model while keeping the model complexity $\Omega(g)$ low (e.g. using the minimum of features). Value (G) is the number of possible explanations, for example, all possible general linear models. The proximity measure πx defines the size of the neighbourhood considered for the explanation around instance x. In practice, LIME only optimises the loss part, and the user has to determine the complexity, for example, by selecting the maximum number of features that the linear regression model may use. Therefore, it is possible to understand whether variables that increase the explanation probability either support or contradict the explanation for a given parcel. Locally, the behaviour of the ML model might be different because the outcomes rely linearly or monotonically on some variables, instead of having a complex dependence on them; therefore, local interpretability might be more accurate than global interpretability (Ribeiro et al., 2016b).

In this study, PFI and PDPs were performed using the 'varim' and 'pdp' packages (Greenwell, 2017; Probst and Janitza, 2020) of R version 4.0.2 statistical software, respectively, while the 'lime' package was used to implement the LIME method (Pedersen and Benesty, 2019).

3. Results

The RF modelling outputs were examined for global interpretation using the PFI method. Fig. 4a–c presents the 10 most influential variables in explaining the land used for wheat, maize, and olive groves. For each cropland, the list of variable importance was distinctive. Although 5 of the 10 variables were the same for the explanation, they presented importance scores quite differently. Overall, the model results indicate that bioclimatic and biophysical variables provided a more significant explanation, while socioeconomic variables were less important and did not seem to affect the model significantly.

The most influential variables (with an importance score of > 50) in explaining land use for wheat plantations were drainage density (V136) and slope (V134), while those for maize plantations were slope, soil type (V132), drainage density, and the ombrothermic index anomaly for humid years (V131). In the case of land used for olive grove plantations, all 10 most influential variables had an importance score of > 50, but the mean minimum temperature of the coldest month (V118), draining density, and ombrothermic index of the summer quarter (V124) were the top three most important.

The PDP method was also used to provide a global interpretation. Fig. 5 presents the response curves for the first six most influential variables and their probability of explaining the land used for wheat plantations. In particular, each plot shows that the probability of land being used for wheat plantations increased on average: (i) by 0.4 as the drainage density increased up to 5 km/km^2, after which the probability did not change; (ii) by 0.4 as the slope increased up to 30%, after which

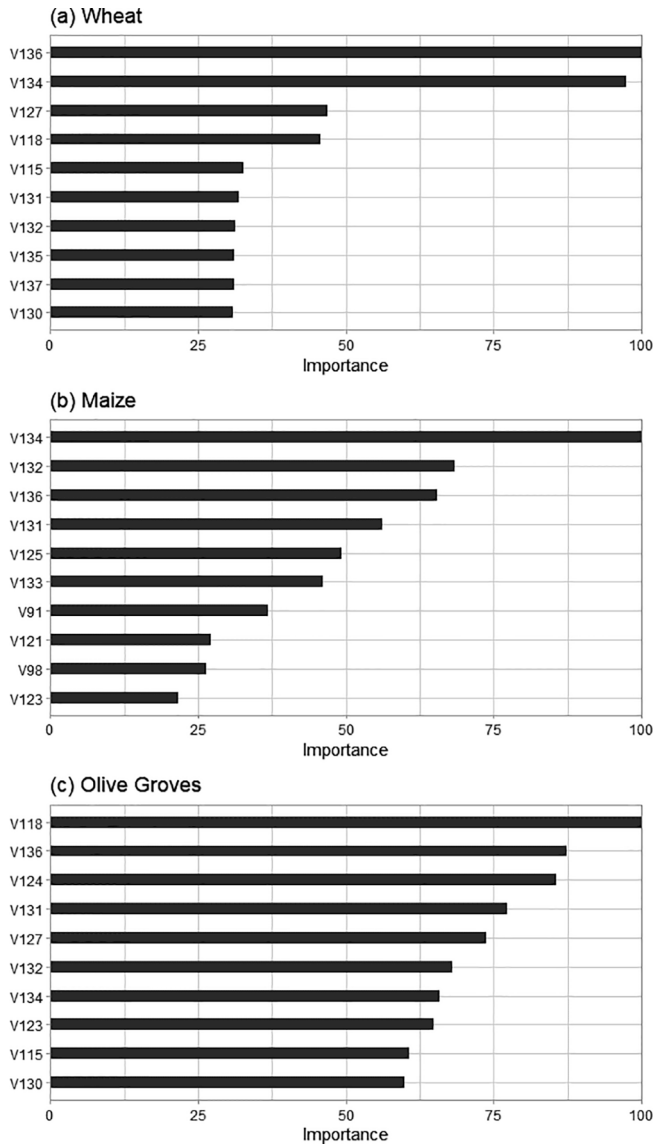


Fig. 4. Ten most influential variables and respective importance score in explaining wheat (a), maize (b), and olive grove (c) plantations. See Table A1 in the Appendix for further details of the variables.

the probability did not change; (iii) by 0.25 as positive precipitation for humid years increased up to 800 mm, after which the probability did not change; (iv) by 0.20 as the mean temperature of the coldest month of the year increased, with a positive impact up to 4.5 °C, whereas the variable effect reduced quickly at > 5 °C before the probability increased again by 0.20 at > 5.5 °C; (v) by 0.18 as the annual positive temperature increased (the variable effect reduced quickly after 1900 °Cx10); and (vi) by 0.16 as the ombrothermic index anomaly for humid years increased up to 0.36, after which the variable affect was negligible.

Fig. 6 displays the response curves for the first six most influential variables and their probability of explaining the land used for maize plantations. In particular, each plot shows that the probability of land being used for maize plantation increased on average: (i) by 0.8 as the slope increase up to 10%, after which the probability did not change; (ii) by 0.5 when the soil type was classified as "A" (incipient soils - modern non-limestone alluvisols of medium texture); (iii) by 0.6 as the drainage density increased up to 8 km/km^2, after which the probability did not change; (iv) by 0.6 as the ombrothermic index anomaly for humid years increased up to 0.34, after which the probability did not change; (v) by 0.4 as the ombrothermic index of the summer quarter plus the previous month increased up to 0.6, after which the variable effect reduced quickly, and the probability increases by 0.2 at > 0.7 °C; and (vi) by 0.3 when the soil capacity was classified as "Ee" (very severe limitations).

Fig. 7 presents the response curves for the first six most influential variables and their threshold value probability of explaining the land used for olive grove plantations. In particular, each plot shows that the probability of land being used for olive grove plantations increased on average: (i) by 0.3 as the mean minimum temperature of the coldest month increases up to 6 °C, after which the probability did not change; (ii) by 0.3 as the draining density increased up to 8 km/km^2, after which the variable affect reduced; (iii) by 0.25 as the ombrothermic index of the summer quarter increased up to 0.3, after which the variable affect reduced quickly; (iv) by 0.16 as the ombrothermic index anomaly for humid years increased up to 0.35, after which the variable affect reduced; (v) by 0.2 as positive precipitation for humid years increased up to 1000 mm, after which the probability did not change; and (vi) by 0.15 when the soil type was classified as "Ex" (incipient soils – lithosols of xeric regime climates, of schist or greywacke).

The LIME method was employed to provide local interpretability (an explanation at the parcel level). Table 2 lists the best probability of four cases for each cropland (each case is a single parcel) and the top five variables that supported or contradicted the local explanation. Although most of the variables supported the explanation for single parcels used for wheat and olive groves, some variables contradicted the explanation (in parcel case #424 for wheat, and parcel cases #2138, #2534, and #2608 for olive groves). For maize, none of the top five variables contradicted this explanation. In general, for each cropland in the analysis, the top five variable sets were similar, although they weighted quite differently.

Fig. 8 displays eight cases, whereby the explanation probability of a parcel was used for wheat plantations. It can be seen that slope (V134), soil type (V132), agricultural holdings with agricultural combine harvester machinery (V53), and drainage density (V136) supported the highest (largest positive weight) explanation probability. In particular, when the slope was < 3.35% and the drainage density was between 2.85 km/km^2 and 3.93 km/km^2, there was a higher probability that a parcel was used for wheat plantation. At the local level, agricultural holdings with agricultural combine harvester machinery increased the probability, specifically > 28 pieces of combine harvester machinery increased the probability of wheat plantations in all eight single parcels.

Fig. 9 presents eight cases and the five most influential variables explaining the parcels used for maize plantations. It can be seen that slope (V134) and drainage density (V136) supported the explanation probability. In particular, when the slope was < 2.24% and the drainage density was < 2.64 km/km^2, there was a higher probability that a parcel was used for maize plantations. In addition, autonomous (legal form) utilised agricultural areas (V91) of ≤ 3213 ha and areas of permanent crops with fresh fruit plantations (V98) of < 86 ha increased the probability of maize plantations in all eight single parcels in the analysis.

Fig. 10 shows eight cases and the most influential variables explaining the parcels used for olive groves plantation. It can be seen that when the mean minimum temperature of the coldest month (V118) was ≤ 4.99 °C and the drainage density (V136) was between 3.09 km/km^2 and 4.10 km/km^2, the probability of a parcel being used for olive grove plantations increased. In addition, the area of permanent crops (V97) of > 5568 ha increased the probability explanation. However, agricultural holdings with < 662 (number) poultry (V6) and ≤ 10% of sole agricultural holders with a level of education outside the agricultural/forestry field (V79) contradicted the explanation. The variables explaining the parcels used for olive groves did not remain constant in all eight parcel cases in the analysis.

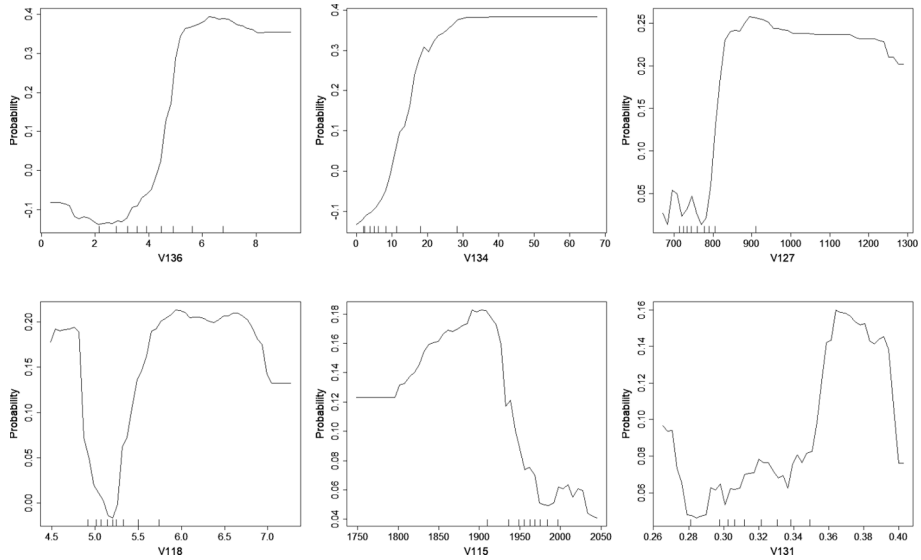


Fig. 5. Response curves for the six most influential variables and their probability of explaining the land used for wheat plantations. See Table A1 in the Appendix for further details of the variables.

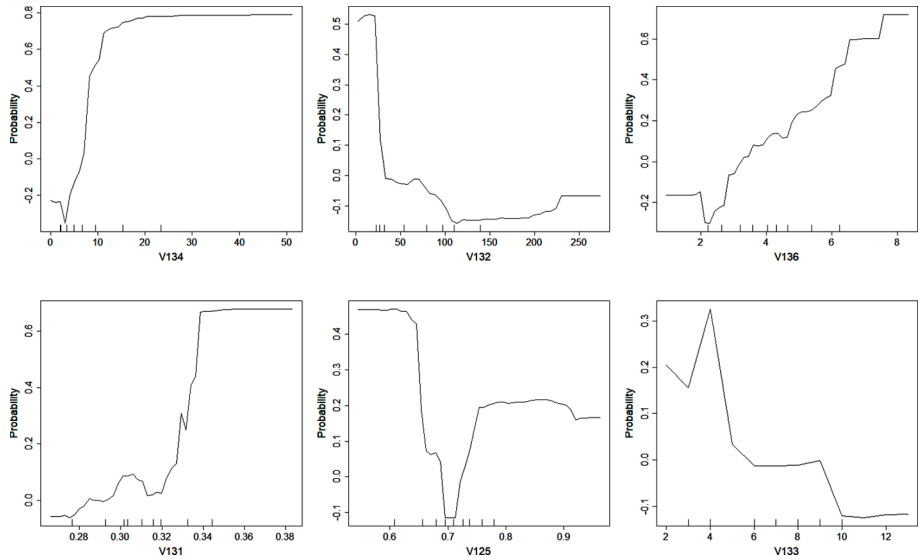


Fig. 6. Response curves for the five most influential variables and their probability of explaining the land used for maize plantations. See Table A1 in the Appendix for further details of the variables.

4. Discussion

4.1. An ML and model-agnostic approach for agricultural land modelling

The modelling outcomes revealed the relationship between the use of agricultural land and environmental and socioeconomic conditions, showing that for each cropland under analysis, the explanatory factors varied significantly. At the global interpretability level, the results showed a highly dominant explanation of drainage density to the land used for wheat, maize, and olive grove plantations. However, the same variable not only exhibited different importance scores for the model interpretability of each cropland, but also presented different threshold values. For instance, in wheat plantations, the probability increased for a drainage density threshold value up to 5 km/km^2, whereas in maize and olive grove plantations, the explanation increased for a threshold up to 8 km/km^2. Overall, the model results emphasise that a high drainage density (>3.5 km/km^2) (Shankar and Mohan, 2006) is an important condition explaining the land used for these three crop plantations. These findings agree with those of other studies, which highlighted the importance of highly drained soils for the rooting depth of crops (Akpoti et al., 2019).

In addition, the slope could increase the explanation regarding wheat and maize plantations; however, the threshold values were substantially different (up to 30% for wheat and 10% for maize). Indeed, the slope of a plantation is a crucial factor for crop growth because it not only affects the vegetation structure but also the internal soil water drainage (Akpoti et al., 2019; Marcos-Martinez et al., 2017). Moreover, the mean minimum temperature of the coldest month explained most of the land used for olive grove plantations. For instance, exposure to cold temperatures is linked to the optimal differentiation of flower buds and the reduction of parasites and pathogens in olive trees (De Melo-Abreu et al., 2004; Rallo and Cuevas, 2017).

At the local interpretability level, the outcomes highlighted that socioeconomic factors became relevant, with differences observed with regard to variable explanation probability. For instance, the degree of mechanisation had a significant probability of supporting the

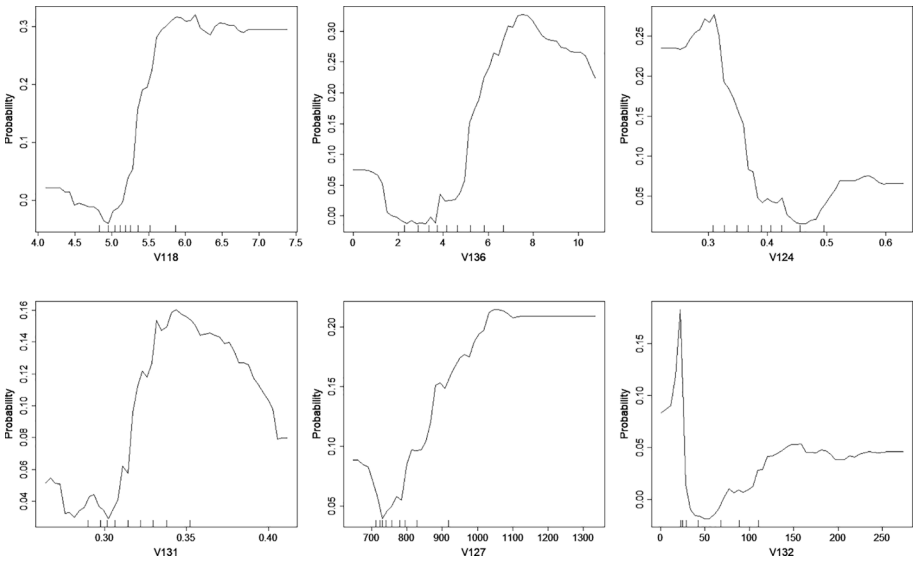


Fig. 7. Response curves for the five most influential variables and their probability of explaining the land used for olive groves. See Table A1 in the Appendix for further details of the variables.

Table 2
Four principal cases of each cropland (the four with the best probability) and the respective five most influential variables that increased (supported) or decreased (contradicted) the explanation. See Table A1 in the Appendix for further details of the variables.

Cropland	Case#	Probability	Top five variables from LIME
Wheat	539	0.99	Supports: V134, V136, V132, V53, V118
Wheat	424	0.98	Supports: V132, V136, V53, V134; Contradicts: V115
Wheat	283	0.97	Supports: V134, V132, V53, V115, V98
Wheat	406	0.96	Supports: V134, V136, V53, V132, V118
Maize	185	1.00	Supports:V134, V98, V91, V132, V131
Maize	152	0.99	Supports:V98, V134, V91, V136, V132
Maize	176	0.99	Supports:V98, V134, V91, V132, V121
Maize	153	0.99	Supports:V134, V98, V91, V132, V136
Olive groves	2138	1.00	Supports: V118, V97, V127, V124; Contradicts: V6
Olive groves	6531	0.99	Supports: V118, V136, V133, V124, V127
Olive groves	2534	0.99	Supports: V118, V136, V124, V95; Contradicts: V79
Olive groves	2608	0.99	Supports: V118, V97, V124, V132; Contradicts: V6

explanation regarding wheat plantation in the eight parcels analysed (Ismail and Abdel-Mageed, 2010). Therefore, the outcome of the analysis indicates that, while environmental factors such as drainage density or slope were important for globally explaining the plantations of the three croplands in the study area, socioeconomic factors became equally important at the parcel level. These findings are consistent with those of other studies (Akpoti et al., 2019; Marcos-Martinez et al., 2017; Santiphop et al., 2012; Thenkabail, 2003).

The results of this study showed that the ML and model-agnostic methods could capture the complex interactions between the human–environmental processes influencing agricultural land use. The identification of the main variables that can explain the use of agricultural land for wheat, maize, and olive grove plantations helps to fill the knowledge gap for modelling these croplands, especially in southern Portugal. Therefore, in geographical regions with conditions similar to those of Mediterranean basins, such factors could be used to better characterise the suitable agricultural areas (Akpoti et al., 2019; Marcos-Martinez et al., 2017).

From a methodological point of view, our study suggests that the developed approach presents a high potential for use as an analytical method in the field of agricultural land systems. For example, the level of interpretability provided by the applied approach provides a reference for land suitability analysis (Akpoti et al., 2019). Certainly, the potential of such an approach deserves further development and testing for other spatiotemporal phenomena of land use to support planning strategies and more efficient and targeted land policies (e.g. research on the driving forces of LULC changes (Aburas et al., 2019)), and/or to anticipate and manage upcoming land changes due to variations in environmental and socioeconomic factors (Baessler and Klotz, 2006; Santiphop et al., 2012).

4.2. Limitations and recommendations

Based on our results and the considerations discussed above, we recognise that the development and implementation of the proposed approach had some limitations. First, some skills and knowledge of statistical software and methods were required. Second, although data availability has increased in many scientific fields, in this study, data related to the multiple factors affecting the use of agricultural land were limited to biophysical, bioclimatic, and socioeconomic factors. As such, the lacking of data related to for example political or cultural factors were a limitation of this study. In fact, having a large amount of data is an important element to strengthen the capability of ML modelling because it ensures adequate training and validation, thereby preventing generalisation problems (Carvalho et al., 2019). However, engaging scientific research with spatially explicit data depends on data availability, which can limit the use of ML models and must be a decisive factor to be considered in subsequent studies using such methods. Third, not having sufficient and representative sample sizes for the three analysed crop plantations can be detrimental to the model-agnostic results. This was, in fact, a second major constraint of this study, and it depicts the limitations associated with such an approach. Moreover, by being a model-agnostic approach, different types of explanations and degrees of interpretability regarding the factors potentially explaining agricultural land use may be obtained (Alvarez-Melis and Jaakkola, 2018; Carvalho et al., 2019; Murdoch et al., 2019; Slack et al., 2020). Therefore, the results should be interpreted critically by field experts who have improved knowledge regarding the underlying functions of agricultural land-use systems. Moreover, our findings need to be used at a regional scale because of their high spatial variability. Fourth, while this study builds upon timely and recent data to provide insights into the

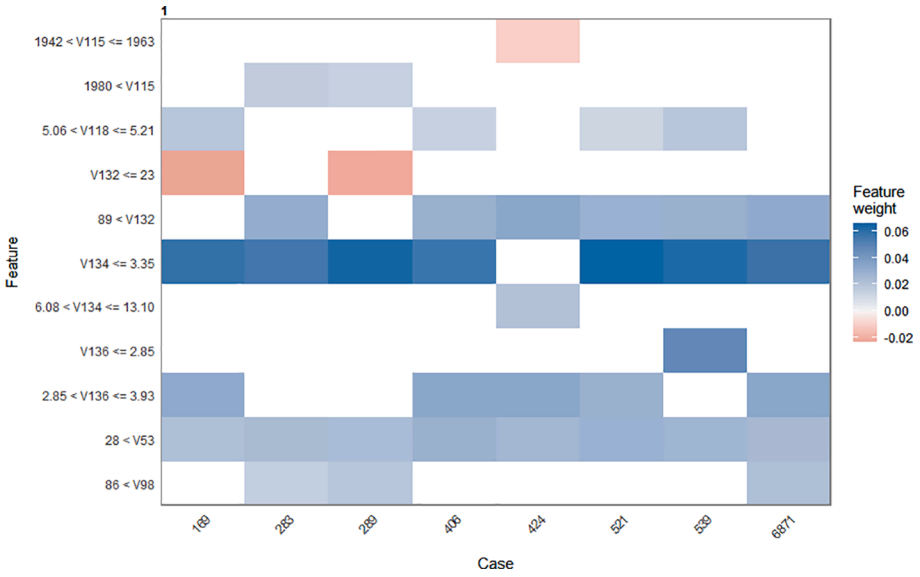


Fig. 8. Most influential variables explaining the parcels used for wheat plantations. See Table A1 in the Appendix for further details of the variables.

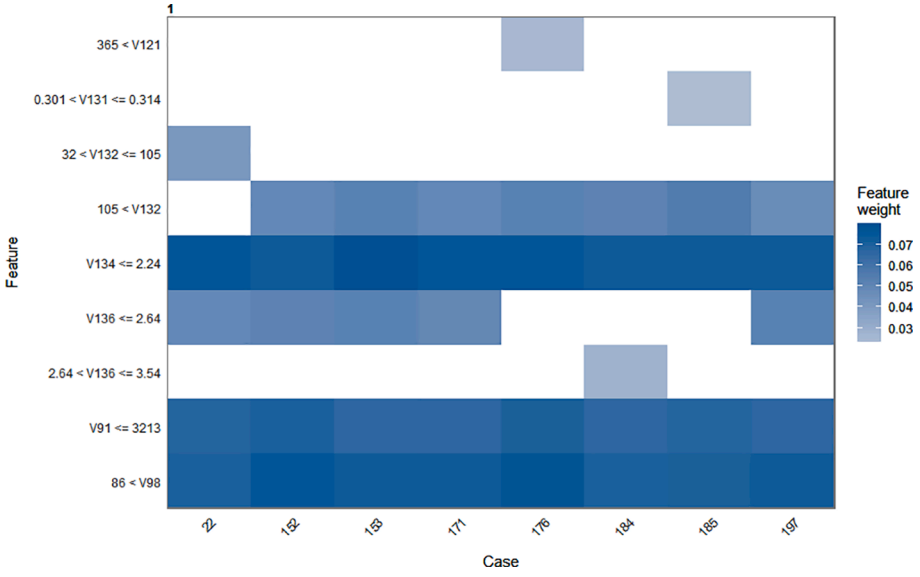


Fig. 9. Most influential variables explaining the parcels used for maize plantations. See Table A1 in the Appendix for further details of the variables.

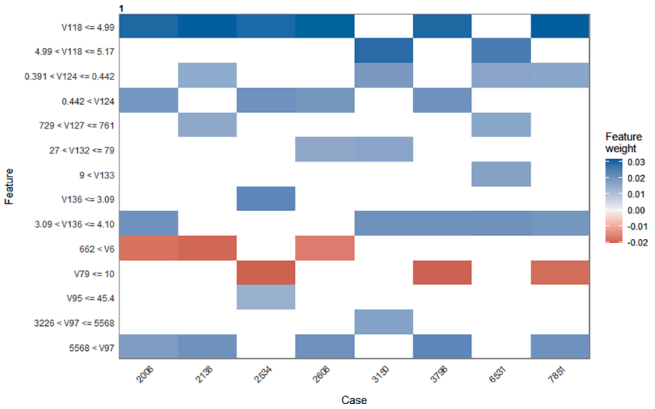


Fig. 10. Most influential variables explaining the parcels used for olive groves plantation. See Table A1 in the Appendix for further details of the variables.

agricultural land used for wheat, maize, and olive grove plantations at a regional scale in a Portuguese district, we acknowledge that more studies should be carried out at different scales and across different geographic contexts to gain a deeper understanding of the underlying factors explaining other important and relevant croplands for global food security and ecosystem services (FAO, 2018). Fifth, and last, we used the RF ML algorithm, PFI, PDPs, and LIME model-agnostic methods, but substantial efforts have already been made in AI and xAI research fields, and different algorithms and methods are readily available (Carvalho et al., 2019; Molnar, 2019). Therefore, future research should focus on comparative studies that could guide new information and improve interpretation (Brun et al., 2020).

5. Conclusions

To comprehensively evaluate the factors that can potentially explain the use of agricultural land for wheat, maize, and olive grove plantations, this study implemented an ML and agnostic-model approach based on agricultural parcel-data sampled points and biophysical,

bioclimatic, and socioeconomic variables. We applied global interpretable methods and identified that drainage density, slope, soil type, and the ombrothermic index anomaly (for humid and dry years) were the five most common important variables explaining the use of agricultural land for the three crop plantations in the study area. Using the local interpretable method, we explored spatial variations and found that socioeconomic variables became relevant at the parcel level. For instance, factors such as the degree of mechanisation could influence specific parcels of wheat. Overall, the analysis outcomes indicated that biophysical and bioclimatic conditions were more influential than socioeconomic conditions. As demonstrated, the proposed analytical approach may be particularly important for research on agricultural land use because it can capture the complex behaviours and underlying functions of agricultural land-use systems, thus providing crucial insights that can help solve several problems related to food production, social stability, and sustainable land use. We believe that this approach is a step towards providing comprehensive assessments of agricultural land use, and has the potential to serve as a decision-making tool to better plan and control the use of agricultural land. Despite the results we have achieved, more studies should be conducted at different scales and across different geographic contexts to gain a deeper understanding. Further research can be used to analyse the underlying factors explaining other important and relevant croplands for global food security and ecosystem services.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Portuguese Foundation for Science and Technology (FCT) under Grant [number SFRH/BD/115497/2016]; Centre for Geographical Studies—Universidade de Lisboa and FCT under Grant [number UIDB/00295/2020 + UIDP/00295/2020]. We would like to thank the GEOMODLAB - Laboratory for Remote Sensing, Geographical Analysis and Modelling—of the Center of Geographical Studies/IGOT for providing the required equipment and software. We would also like to thank to the editor and the anonymous reviewers that contributed to the improvement of this paper.

Funding

This work was supported by the Portuguese Foundation for Science and Technology (FCT) under Grant [number SFRH/BD/115497/2016]; Centre for Geographical Studies—Universidade de Lisboa and FCT under Grant [number UIDB/00295/2020 + UIDP/00295/2020].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2021.108200.

References

Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience Remote Sens. 57, 1–20. https://doi.org/10.1080/15481603.2019.1650447.

Aburas, M.M., Ahamad, M.S.S., Omar, N.Q., 2019. Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: a review. Environ. Monit. Assess. 191, 1–28. https://doi.org/10.1007/s10661-019-7330-6.

Akpoti, K., Kabo-bah, A.T., Zwart, S.J., 2019. Agricultural land suitability analysis: State-of-the-art and outlooks for integration of climate change analysis. Agric. Syst. https://doi.org/10.1016/j.agsy.2019.02.013.

Al-Fugara, A., Pourghasemi, H.R., Al-Shabeeb, A.R., Habib, M., Al-Adamat, R., Al-Amoush, H., Collins, A.L., 2020. A comparison of machine learning models for the mapping of groundwater spring potential. Environ. Earth Sci. 2020 7910 79, 1–19. https://doi.org/10.1007/S12665-020-08944-1.

Alvarez-Melis, D., Jaakkola, T.S., 2018. On the Robustness of Interpretability Methods, in: ICML Workshop on Human Interpretability in Machine Learning (WHI 2018). Stockholm, Sweden.

Apley, D.W., Zhu, J., 2016. Visualizing the effects of predictor variables in black box supervised learning models. J. R Stat. Soc. Ser. B Stat. Methodol. 82, 1059–1086.

Baessler, C., Klotz, S., 2006. Effects of changes in agricultural land-use on landscape structure and arable weed vegetation over the last 50 years. Agric. Ecosyst. Environ. 115, 43–50. https://doi.org/10.1016/j.agee.2005.12.007.

Braimoh, A.K., 2009. Agricultural land-use change during economic reforms in Ghana. Land Policy 26, 763–771. https://doi.org/10.1016/j.landusepol.2008.10.006.

Brun, P., Thuiller, W., Chauvier, Y., Pellissier, L., Wüest, R.O., Wang, Z., Zimmermann, N. E., 2020. Model complexity affects species distribution projections under climate change. J. Biogeogr. 47, 130–142. https://doi.org/10.1111/JBI.13734.

Cartone, A., Postiglione, P., 2020. Principal component analysis for geographical data: the role of spatial effects in the definition of composite indicators. Spat. Econ. Anal. 1–22 https://doi.org/10.1080/17421772.2020.1775876.

Carvalho, D.V., Pereira, E.M., Cardoso, J.S., 2019. Machine learning interpretability: a survey on methods and metrics. Electronics 8, 832. https://doi.org/10.3390/electronics8080832.

Castillo, C.P., Jacobs-Crisioni, C., Diogo, V., Lavalle, C., 2021. Modelling agricultural land abandonment in a fine spatial resolution multi-level land-use model: an application for the EU. Environ. Model. Softw. 136, 104946 https://doi.org/10.1016/j.envsoft.2020.104946.

Correia, T.P., 1993. Threatened landscape in Alentejo, Portugal: the 'montado' and other 'agro-silvo-pastoral' systems. Landsc. Urban Plan. 24, 43–48. https://doi.org/10.1016/0169-2046(93)90081-N.

De Melo-Abreu, J.P., Barranco, D., Cordeiro, A.M., Tous, J., Rogado, B.M., Villalobos, F. J., 2004. Modelling olive flowering date using chilling for dormancy release and thermal time. Agric. For. Meteorol. 125, 117–127. https://doi.org/10.1016/j.agrformet.2004.02.009.

Demšar, U., Harris, P., Brunsdon, C., Fotheringham, A.S., McLoone, S., 2013. Principal component analysis on spatial data: an overview. Ann. Assoc. Am. Geogr. 103, 106–128. https://doi.org/10.1080/00045608.2012.689236.

DGT, 2018. Especificações Técnicas da Carta de Uso e Ocupação do Solo (COS) de Portugal Continental para 1995, 2007, 2010 e 2015. Lisboa.

Dohoo, I.R., Ducrot, C., Fourichon, C., Donald, A., Hurnik, D., 1997. An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. Prev. Vet. Med. 29, 221–239. https://doi.org/10.1016/S0167-5877(96)01074-4.

FAO, 2018. Crop Prospects and Food Situation. Italy, Rome.

FAO, 2017. The Future of Food and Agriculture - Trends and Challenges. Rome.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D., Zaks, D.P.M., 2011. Solutions for a cultivated planet. Nature 478, 337–342. https://doi.org/10.1038/nature10452.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food security: the challenge of feeding 9 billion people. Science 80-. https://doi.org/10.1126/science.1185383.

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2013. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J. Comput. Graph. Stat. 24, 44–65.

Gomes, E., Banos, A., Abrantes, P., Rocha, J., Kristensen, S.B.P., Busck, A., 2019. Agricultural land fragmentation analysis in a peri-urban context: from the past into the future. Ecol. Indic. 97, 380–388. https://doi.org/10.1016/j.ecolind.2018.10.025.

Greenwell, B., 2017. pdp: an R package for constructing partial dependence plots. R J. 9, 421–436.

Hagenauer, J., Omrani, H., Helbich, M., 2019. Assessing the performance of 38 machine learning models: the case of land consumption rates in Bavaria, Germany. Int. J. Geogr. Inf. Sci. 33, 1399–1419. https://doi.org/10.1080/13658816.2019.1579333.

INE, 2012. Censos 2011 Resultados Definitivos – Região Alentejo. Instituto Nacional de Estatística, Lisboa: Instituto Nacional de Estatística.

Ismail, Z.E., Abdel-Mageed, A.E., 2010. WORKABILITY AND MACHINERY PERFORMANCE FOR WHEAT HARVESTING. Misr J. Agric. Eng. 27, 90–103. https://doi.org/10.21608/mjae.2010.106860.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to Statistical Learning: With Applications in R., Corr. 7th printing. ed, Current medicinal chemistry. Springer. https://doi.org/10.1007/978-1-4614-7138-7.

Johnston, R., Jones, K., Manley, D., 2018. Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. Qual. Quant. 52, 1957–1976. https://doi.org/10.1007/s11135-017-0584-6.

Jombart, T., Devillard, S., Dufour, A.B., Pontier, D., 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity (Edinb) 101, 92–103. https://doi.org/10.1038/hdy.2008.34.

Koohafkan, P. (Parviz), Altieri, M.A., 2016. Forgotten agricultural heritage : reconnecting food systems and sustainable development.

Kourgialas, N.N., 2021. A critical review of water resources in Greece: the key role of agricultural adaptation to climate-water effects. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2021.145857.

Lambin, E.F., Turner, B.L., Geist, H.J., Agbola, S.B., Angelsen, A., Folke, C., Bruce, J.W., Coomes, O.T., Dirzo, R., George, P.S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E.F., Mortimore, M., Ramakrishnan, P.S., Richards, J.F., Steffen, W., Stone, G.D., Svedin, U., Veldkamp, T.A., 2001. The causes of land-use and land-cover change : moving beyond the myths 11, 261–269.

Lehikoinen, A., Olsson, J., Bergström, L., Bergström, U., Bryhn, A., Fredriksson, R., Uusitalo, L., 2019. Evaluating complex relationships between ecological indicators and environmental factors in the Baltic Sea: a machine learning approach. Ecol. Indic. 101, 117–125. https://doi.org/10.1016/j.ecolind.2018.12.053.

Li, S., Juhász-Horváth, L., Pintér, L., Rounsevell, M.D.A., Harrison, P.A., 2018. Modelling regional cropping patterns under scenarios of climate and socio-economic change in Hungary. Sci. Total Environ. 622–623, 1611–1620. https://doi.org/10.1016/j.scitotenv.2017.10.038.

Li, X., Chen, W., Cheng, X., Wang, L., 2016. A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery. Remote Sens. https://doi.org/10.3390/rs8060514.

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2, 18–22.

Lima, A.R., Cannon, A.J., Hsieh, W.W., 2015. Nonlinear regression in environmental sciences using extreme learning machines: a comparative evaluation. Environ. Model. Softw. 73, 175–188. https://doi.org/10.1016/j.envsoft.2015.08.002.

Lin, F.J., 2008. Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. Qual. Quant. 42, 417–426. https://doi.org/10.1007/s11135-006-9055-1.

Loumou, A., Giourga, C., 2003. Olive groves: "The life and identity of the Mediterranean". Agric. Human Values 20, 87–95. https://doi.org/10.1023/A:1022444005336.

Luan, J., Zhang, C., Xu, B., Xue, Y., Ren, Y., 2020. The predictive performances of random forest models with limited sample size and different species traits. Fish. Res. 227, 105534 https://doi.org/10.1016/J.FISHRES.2020.105534.

Marcos-Martinez, R., Bryan, B.A., Connor, J.D., King, D., 2017. Agricultural land-use dynamics: assessing the relative importance of socioeconomic and biophysical drivers for more targeted policy. Land Policy 63, 53–66. https://doi.org/10.1016/j.landusepol.2017.01.011.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model. Softw. 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001.

Moghaddam, D.D., Rahmati, O., Panahi, M., Tiefenbacher, J., Darabi, H., Haghizadeh, A., Haghighi, A.T., Nalivan, O.A., Tien Bui, D., 2020. The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers. Catena 187, 104421. https://doi.org/10.1016/J.CATENA.2019.104421.

Molnar, C., 2019. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. Available online: https://christophm.github.io/interpretable-ml-book/ (accessed on 22 January 2021).

Monteiro-Henriques, T., Martins, M.J., Cerdeira, J.O., Silva, P., Arsénio, P., Silva, Á., Bellu, A., Costa, J.C., 2016. Bioclimatological mapping tackling uncertainty propagation: application to mainland Portugal. Int. J. Climatol. 36, 400–411. https://doi.org/10.1002/joc.4357.

Muñoz-Rojas, J., Pinto-Correia, T., Hvarregaard Thorsoe, M., Noe, E., 2019. The Portuguese Montado : A Complex System under Tension between Different Land Use Management Paradigms , in: Silvicultures - Management and Conservation. IntechOpen. https://doi.org/10.5772/intechopen.86102.

Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B., 2019. Definitions, methods, and applications in interpretable machine learning. Proc. Natl. Acad. Sci. U.S.A. 116, 22071–22080. https://doi.org/10.1073/pnas.1900654116.

Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., Toxopeus, A.G., 2014. Where is positional uncertainty a problem for species distribution modelling? Ecography (Cop.) 37, 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x.

Ndamani, F., Watanabe, T., 2017. Developing indicators for adaptation decision-making under climate change in agriculture: a proposed evaluation model. Ecol. Indic. 76, 366–375. https://doi.org/10.1016/j.ecolind.2016.12.012.

Pedersen, T.L., Benesty, M., 2019. lime: Local interpretable model-agnostic explanations.

Petit, C., Aubry, C., Rémy-Hall, E., 2011. Agriculture and proximity to roads: How should farmers and retailers adapt? Examples from the Ile-de-France region. Land Policy 28, 867–876. https://doi.org/10.1016/j.landusepol.2011.03.001.

Portugal, S., 2019. Recenseamento Agrícola. [WWW Document]. URL https://ra09.ine.pt/.

Postek, P., Leń, P., Stręk, Ż., 2019. The proposed indicator of fragmentation of agricultural land. Ecol. Indic. 103, 581–588. https://doi.org/10.1016/j.ecolind.2019.04.023.

Probst, P., Boulesteix, A.-L., 2018. To tune or not to tune the number of trees in random forest. J. Mach. Learn. Res. 10, 3242038. https://doi.org/10.5555/3122009.3242038.

Probst, P., Janitza, S., 2020. varImp: RF Variable Importance for Arbitrary Measures.

Qi, Y., 2012. Random forest for bioinformatics. Ensemble Mach. Learn. 307–323 https://doi.org/10.1007/978-1-4419-9326-7_11.

Raczko, E., Zagajewski, B., 2017. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. Eur. J. Remote Sens. 50, 144–154. https://doi.org/10.1080/22797254.2017.1299557.

Radwan, T.M., Blackburn, G.A., Whyatt, J.D., Atkinson, P.M., 2019. Dramatic loss of agricultural land due to urban expansion threatens food security in the Nile delta, Egypt. Remote Sens. 11, 332. https://doi.org/10.3390/rs11030332.

Rallo, L., Cuevas, J., 2017. Fructificación y producción (Fruiting and production). In: Barranco, D., Fernandez-Escobar, R., Rallo, L. (Eds.), El Cultivo Del Olivo (Olive Growing). Mundi-Prensa, Madrid, pp. 145–186.

Ren, X., Mi, Z., Georgopoulos, P.G., 2020. Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution: modeling ozone concentrations across the contiguous United States. Environ. Int. 142, 105827 https://doi.org/10.1016/j.envint.2020.105827.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016a. Model-Agnostic Interpretability of Machine Learning, in: ICML Workshop on Human Interpretability in Machine Learning.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. "Why should i trust you?" Explaining the predictions of any classifier, in. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, pp. 1135–1144. https://doi.org/10.1145/2939672.2939778.

Salami, D., Sousa, C.A., Martins, M., do Capinha, R.O.C., 2020. Predicting dengue importation into Europe, using machine learning and model-agnostic methods. Sci. Rep. 10, 1–13. https://doi.org/10.1038/s41598-020-66650-1.

Santiphop, T., Shrestha, R.P., Hazarika, M.K., 2012. An analysis of factors affecting agricultural land use patterns and livelihood strategies of farm households in Kanchanaburi Province, Thailand. J. Land Use Sci. 7, 331–348. https://doi.org/10.1080/1747423X.2011.587208.

Shankar, M.N.R., Mohan, G., 2006. Assessment of the groundwater potential and quality in Bhatsa and Kalu river basins of Thane district, western Deccan Volcanic Province of India. Environ. Geol. 49, 990–998. https://doi.org/10.1007/s00254-005-0137-5.

Shortridge, J.E., Guikema, S.D., Zaitchik, B.F., 2016. Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrol. Earth Syst. Sci. 20, 2611–2628. https://doi.org/10.5194/hess-20-2611-2016.

Slack, D., Hilgard, S., Jia, E., Singh, S., Lakkaraju, H., 2019. Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods. AIES 2020 - Proc. AAAI/ACM Conf. AI, Ethics, Soc. 180–186.

Thenkabail, P.S., 2003. Biophysical and yield information for precision farming from near-real-time and historical Landsat TM images. Int. J. Remote Sens. 24, 2879–2904. https://doi.org/10.1080/01431160710155974.

Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D.J., Franz, K.J., 2020. Mapping of 30-meter resolution tile-drained croplands using a geospatial modeling approach. Sci. Data 7, 1–10. https://doi.org/10.1038/s41597-020-00596-x.

Velásquez-Milla, D., Casas, A., Torres-Guevara, J., Cruz-Soriano, A., 2011. Ecological and socio-cultural factors influencing in situ conservation of crop diversity by traditional Andean households in Peru. J. Ethnobiol. Ethnomed. 7, 40. https://doi.org/10.1186/1746-4269-7-40.

Viana, C.M., Girão, I., Rocha, J., 2019. Long-term satellite image time-series for land use/land cover change detection using refined open source data in a rural region. Remote Sens. 11, 1104. https://doi.org/10.3390/rs11091104.

Viana, C.M., Rocha, J., 2020. Evaluating dominant land use/land cover changes and predicting future scenario in a rural region using a memoryless stochastic method. Sustainability 12, 4332. https://doi.org/10.3390/su12104332.

Winkler, A.M., Webster, M.A., Vidaurre, D., Nichols, T.E., Smith, S.M., 2015. Multi-level block permutation. Neuroimage 123, 253–268. https://doi.org/10.1016/j.neuroimage.2015.05.092.

Wu, W. Bin, Yu, Q.Y., Peter, V.H., You, L.Z., Yang, P., Tang, H.J., 2014. How could agricultural land systems contribute to raise food production under global change? J. Integr. Agric. https://doi.org/10.1016/S2095-3119(14)60819-4.

Wu, L., Zhu, X., Lawes, R., Dunkerley, D., Zhang, H., 2019. Comparison of machine learning algorithms for classification of LiDAR points for characterization of canola canopy structure. Int. J. Remote Sens. 40, 5973–5991. https://doi.org/10.1080/01431161.2019.1584929.

Yang, R.M., Zhang, G.L., Liu, F., Lu, Y.Y., Yang, Fan, Yang, Fei, Yang, M., Zhao, Y.G., Li, D.C., 2016. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. Ecol. Indic. 60, 870–878. https://doi.org/10.1016/J.ECOLIND.2015.08.036.

Zhang, Z., Xu, S., Capinha, C., Weterings, R., Gao, T., 2019. Using species distribution model to predict the impact of climate change on the potential distribution of Japanese whiting Sillago japonica. Ecol. Indic. 104, 333–340. https://doi.org/10.1016/j.ecolind.2019.05.023.